Iterative Translation Refinement with Large Language Models

Pinzhen Chen\textsuperscript{1} \hspace{1cm} Zhicheng Guo\textsuperscript{2} \hspace{1cm} Barry Haddow\textsuperscript{1} \hspace{1cm} Kenneth Heafield\textsuperscript{1}

\textsuperscript{1}School of Informatics, University of Edinburgh  
\textsuperscript{2}Dept. of Comp. Sci. & Tech., Institute for AI, Tsinghua University  
{pinzhen.chen,bhaddow,kenneth.heafield}@ed.ac.uk  
guo-zc21@mails.tsinghua.edu.cn

Abstract

Large language models have shown surprising performances in understanding instructions and performing natural language tasks. In this paper, we propose iterative translation refinement to leverage the power of large language models for more natural translation and post-editing. We show that by simply involving a large language model in an iterative process, the output quality improves beyond mere translation. Extensive test scenarios with GPT-3.5 reveal that although iterations reduce string-based metric scores, neural metrics indicate comparable if not improved translation quality. Further, human evaluations demonstrate that our method effectively reduces “translationese” compared to initial GPT translations and even human references, especially for into-English directions. Ablation studies underscore the importance of anchoring the refinement process to the source input and a reasonable initial translation.

1 Introduction

Large language models (LLMs), e.g. generative pre-trained Transformers (GPT), have made significant advancements across a range of natural language processing tasks. (Radford et al., 2019; Kaplan et al., 2020; Brown et al., 2020; Chowdhery et al., 2022; Ouyang et al., 2022). In machine translation, where the convention is to use an encoder-decoder architecture to deal with source and target sentences directly (Bahdanau et al., 2015; Vaswani et al., 2017), recent studies explored the feasibility of prompting LLMs for translation (Vilar et al., 2022; Zhang et al., 2023). GPT yields high-quality translations for high-resource languages (Hendy et al., 2023), achieves remarkable results on quality estimation (Kocmi and Federmann, 2023), and produces less literal translations especially out-of-English (Raunak et al., 2023a). Very recently, it is shown that post-editing leveraging GPT-4 achieves promising results (Raunak et al., 2023b).

Previous LLM research does not extensively explore the “translationese” phenomenon, which refers to an unnatural translation due to both source language interference and the translation process itself (Gellerstam, 1986; Baker, 1996; Teich, 2003). It is undesirable but appears in various stages: data (Riley et al., 2020), model outputs (Freitag et al., 2020a), human post-edited translations (“post-editese”, Toral, 2019), and human references (Freitag et al., 2020b). Even LLMs might be translationese-prone as their translation power is associated with implicit bilingual signals (Briakou et al., 2023). We hence propose a simple way to obtain higher-quality translations from LLMs: iterative translation refinement. Building on automatic post-editing which imitates human corrections (APE, Knight and Chander, 1994; Chatterjee et al., 2018), we prompt GPT for a translation and then provide the source-translation pair to query for a refined translation in multiple rounds.

Our method offers two strengths for combating translationese: 1) LLM prompting allows for iterative and arbitrary re-writing compared to APE which is limited to error fixing without style improvement (Ive et al., 2020); 2) incorporating natural language data leads to more natural translations (Sennrich et al., 2016; Freitag et al., 2019), and LLMs have seen target-side data orders of magnitude larger than datasets for translation or post-editing. Empirical results indicate that the refinement process introduces significant textual changes, but attains higher or similar neural metric scores compared to initial translations. Native speakers prefer the refined outputs in terms of reduced translationese, which is significantly more prevalent in initial GPT translations and human references. The improvements are made without directly optimizing towards translation or post-editing, and as corroborated by recent studies, are challenging to capture through automatic evaluation alone (Freitag et al., 2019, 2022).
Table 1: Prompts used in our work, where \( \{\text{variable}\} \) is substituted with its corresponding content.

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Source: ( {\text{source}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translate</td>
<td>Please give me a translation in ( {\text{lang}} ) without any explanation.</td>
</tr>
<tr>
<td>Refine</td>
<td>Source: ( {\text{source}} )</td>
</tr>
<tr>
<td></td>
<td>Translation: ( {\text{prev_translation}} )</td>
</tr>
<tr>
<td></td>
<td>Please give me a better ( {\text{lang}} ) translation without any explanation.</td>
</tr>
<tr>
<td>RefineContrast</td>
<td>Source: ( {\text{source}} )</td>
</tr>
<tr>
<td></td>
<td>Bad translation: ( {\text{prev_translation}} )</td>
</tr>
<tr>
<td></td>
<td>Please give me a better ( {\text{lang}} ) translation without any explanation.</td>
</tr>
<tr>
<td>RefineRandom</td>
<td>Source: ( {\text{source}} )</td>
</tr>
<tr>
<td></td>
<td>Bad translation: ( {\text{random_target}} ) if first-round, else ( {\text{prev_translation}} )</td>
</tr>
<tr>
<td></td>
<td>Please give me a better ( {\text{lang}} ) translation without any explanation.</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>Sentence: ( {\text{prev_translation}} )</td>
</tr>
<tr>
<td></td>
<td>Please give me a paraphrase in ( {\text{lang}} ) without any explanation.</td>
</tr>
</tbody>
</table>

2 Methodology

2.1 Machine translation and post-editing

Having an input source sentence \( x \) and an optimizable model \( \theta_{\text{mt}} \), the process to obtain a translation \( y \) can be modelled as \( y = \arg\max_y P(y|x, \theta_{\text{mt}}) \). Next, an automatic post-editor \( \theta_{\text{ape}} \) creates a refined translation \( y' \) through \( y' = \arg\max_{y'} P(y'|x, y, \theta_{\text{ape}}) \). Conventional translation or automatic post-editing models train on \( (x, y) \) or \( (x, y, y') \) data pairs.

2.2 LLM Prompting

Our work uses a zero-shot hard prompting strategy by affixing a task description to the input to form a prompt \( p \), and querying an LLM \( \theta_{\text{LLM}} \) to elicit a response, which is treated as the output (Brown et al., 2020). We introduce five prompts in our study:

1. **Translate**: we use this to query for a translation of a source input, extending the above translation process with a prompt \( p \): \( y = \arg\max_y P(y|p, \theta_{\text{LLM}}) \).

2. **Refine**: similar to APE, the LLM is given the source sentence and the previous translation to produce a better translation \( y' = \arg\max_{y'} P(y'|p, y, \theta_{\text{LLM}}) \).

3. **RefineContrast**: as a contrastive prompt to the above, we insert the word “bad” to hint that the previous translation is of low quality, regardless of its actual quality.

4. **RefineRandom**: same prompt text as **RefineContrast**, but in the first iteration, a random target sentence is fed instead of a translation.

5. **Paraphrase**: as an ablation investigation, we prompt to rephrase a translation without seeing the source input \( x \): \( y'' = \arg\max_{y''} P(y''|p, y, \theta_{\text{LLM}}) \). Our work also proposes to iteratively call the refinement and paraphrase prompts, where the source sentence stays the same but the previous translation is updated with the current translation to study how returned translations change. To ensure a parsable high-quality response, in all prompts, we ask the model not to give any explanation. The prompt texts we use are displayed in Table 1.

3 Experiments

3.1 Data, model, and evaluation

We experiment with language pairs from the translation shared task hosted at WMT 2021 and 2022 (Akhbardeh et al., 2021; Kocmi et al., 2022). Language pairs are chosen based on two criteria: 1) the references should not have been seen by GPT, and 2) the target language should have appeared in previous years’ WMT translation tasks, so that COMET has been fine-tuned to provide reliable scores. Specific languages are introduced in the sections later on. We directly benchmark on the test sets, and when multiple references are available, we use human reference “A” released by the organizers as our reference.

The LLM we experiment with is GPT-3.5, the current most powerful API from OpenAI that can be accessed by all users.\(^1\) We use the chat mode as

\(\text{https://platform.openai.com/docs/models/gpt-3-5}\). We use gpt-3.5-turbo which has training data up to Sep 2021, so it should not have seen WMT 2021 or 2022 test references. Nevertheless, our findings are mostly drawn from reference-less and human evaluation.

\(\text{https://platform.openai.com/docs/models/gpt-3-5}\). We use gpt-3.5-turbo which has training data up to Sep 2021, so it should not have seen WMT 2021 or 2022 test references. Nevertheless, our findings are mostly drawn from reference-less and human evaluation.
opposed to the completion mode. As the OpenAI API is very slow to query, for each language pair, we randomly sample 200 instances from the official test set to form our own test. In the refinement and paraphrase experiments, we use the response from the Translate query as the base translation to be improved upon. Similar to the black-box condition in APE, we do not keep the query history to prevent the model from knowing that the previous translation is produced by itself. Overall, the iterated experiments are done four times.

For evaluation, we use automatic text metrics: BLEU (Papineni et al., 2002) and chrF++ (Popović, 2017) as implemented in the sacreBLEU toolkit (Post, 2018).

We also use neural metrics which are proven to better correlate with human judgments (Freitag et al., 2022): COMETDA and COMETQE (Rei et al., 2020). The former requires a reference and the latter is reference-free.

3.2 High-resource experiments with GPT

Our main experiments are conducted on two high-resource language pairs and in total four directions: English-German (en↔de) and English-Chinese (en↔zh) in the WMT 2021 news translation task. We display the automatic scores for all prompts and human references in Table 2. For refinement and paraphrasing experiments, the best iteration is picked according to COMETQE.

We observe that the refined translations achieve COMET scores comparable to or higher than the initial GPT translation, but with a drastic drop in text metrics, indicating huge lexical and structural variations. In terms of COMETQE, refined outputs surpass all initial GPT translations and three out of four human references, with substantial improvement for into-English directions. Paraphrase sees declined numbers in all metrics. This suggests the importance of feeding the source sentence as an anchor during refinement to maintain quality.

Table 2: Automatic scores of different strategies with GPT on high-resource pairs from WMT 2021 news translation.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>BLEU</th>
<th>chrF++</th>
<th>COMETDA</th>
<th>COMETQE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translate</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.0919</td>
</tr>
<tr>
<td>Refine</td>
<td>30.90</td>
<td>57.55</td>
<td>.8066</td>
<td>.1128</td>
</tr>
<tr>
<td>RefineContrast</td>
<td>23.14</td>
<td>51.91</td>
<td>.8525</td>
<td>.1116</td>
</tr>
<tr>
<td>RefineRandom</td>
<td>22.88</td>
<td>52.47</td>
<td>.8452</td>
<td>.1162</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>18.83</td>
<td>51.79</td>
<td>.7777</td>
<td>.0770</td>
</tr>
<tr>
<td>ReferenceA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.1127</td>
</tr>
<tr>
<td>Reference</td>
<td>25.39</td>
<td>53.54</td>
<td>.8427</td>
<td>.1083</td>
</tr>
<tr>
<td>Refine</td>
<td>22.35</td>
<td>50.57</td>
<td>.8478</td>
<td>.1153</td>
</tr>
<tr>
<td>RefineContrast</td>
<td>22.54</td>
<td>51.21</td>
<td>.8211</td>
<td>.0929</td>
</tr>
<tr>
<td>RefineRandom</td>
<td>19.36</td>
<td>46.56</td>
<td>.7906</td>
<td>.0832</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>13.60</td>
<td>43.54</td>
<td>.8197</td>
<td>.1006</td>
</tr>
<tr>
<td>ReferenceA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.0708</td>
</tr>
<tr>
<td>Reference</td>
<td>25.64</td>
<td>53.74</td>
<td>.8199</td>
<td>.0867</td>
</tr>
<tr>
<td>Refine</td>
<td>20.26</td>
<td>49.06</td>
<td>.8156</td>
<td>.0921</td>
</tr>
<tr>
<td>RefineContrast</td>
<td>24.81</td>
<td>51.77</td>
<td>.8538</td>
<td>.1132</td>
</tr>
<tr>
<td>RefineRandom</td>
<td>24.24</td>
<td>47.11</td>
<td>.8323</td>
<td>.1022</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>12.76</td>
<td>40.92</td>
<td>.7931</td>
<td>.0885</td>
</tr>
<tr>
<td>ReferenceA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.0956</td>
</tr>
<tr>
<td>Reference</td>
<td>29.28</td>
<td>20.61</td>
<td>.8300</td>
<td>.0761</td>
</tr>
<tr>
<td>Refine</td>
<td>28.25</td>
<td>19.28</td>
<td>.8417</td>
<td>.0870</td>
</tr>
<tr>
<td>RefineContrast</td>
<td>29.28</td>
<td>19.69</td>
<td>.8395</td>
<td>.0881</td>
</tr>
<tr>
<td>RefineRandom</td>
<td>25.71</td>
<td>17.49</td>
<td>.8126</td>
<td>.0763</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>21.95</td>
<td>17.14</td>
<td>.8144</td>
<td>.0716</td>
</tr>
</tbody>
</table>

Table 2: Automatic scores of different strategies with GPT on high-resource pairs from WMT 2021 news translation.

3.3 Human evaluation on translationese

Volatile textual changes without neural metric degradation could reflect a change in the degree of translationese (Freitag et al., 2020b). Hence, we conduct human evaluations on translation outputs, to assess which translation strategies are more translationese-prone without reference to the source texts. Since the term “translationese” is not commonly known, we mimic Lembersky et al. (2012)’s method by presenting a native speaker with two translations and ask: Please choose the translation that is more fluent, natural, and reflecting better use of ${language}.

We conduct such pairwise evaluation to com-
Figure 1: BLEU, COMET\textsubscript{DA}, and COMET\textsubscript{QE} at different refinement and paraphrase iterations.

3.4 Experiments with WMT submissions

In addition to translation refinement with GPT itself, we also apply our technique to outputs from conventional machine translation models and human translations. These translations can represent genuine errors, if any, introduced during the translation process. We use WMT 2021 German-to-English submissions in the news translation shared task as a starting point.

We select outputs from four models built by research labs that, based on human evaluation, have been ranked at significantly different positions on the German-to-English leaderboard: Tencent (Wang et al., 2021), Facebook AI (Tran et al., 2021), Edinburgh (Chen et al., 2021), and Huawei TSC (Wei et al., 2021). These are competitive systems built with data augmentation, multilingualism,
Finally human reference “B” is added to complement our test on human translations.\footnote{The description paper of WMT 2021 states that “for German→English, the ‘B’ reference was found to be a post-edited version of one of the participating online systems”. We discover that it refers to English→German only, and German→English is not affected.}

A pattern similar to previous results is noticed. For five out of seven submissions, the refinement strategy reaches a higher COMET\textsubscript{QE} score, surprisingly, with up to one-third drop in BLEU. \texttt{Refine\textsubscript{Contrast}} wins \texttt{Refine} notably. \texttt{Paraphrase} records the lowest scores compared to the original submissions and refinements.

### 3.5 Low-resource experiments with GPT

Finally, we run experiments on low-resource and medium-resource translation directions from WMT 2022: English-to-Japanese (en→ja), German-to-French (de→fr), Yakut-to-Russian (sah→ru), and Ukrainian-to-Czech (uk→cs). As Table 5 displays, \texttt{Refine} works the best, obtaining higher COMET\textsubscript{QE} than GPT translations and \texttt{Refine\textsubscript{Contrast}}, except for sah→ru. However, we suspect that sah→ru COMET numbers are less meaningful because the Yakut language is not supported by COMET. Through the 4 BLEU and chrF++ drop in de→fr and the single digit change in lower-resourced en→ja and uk→cs, we find that, during refinement, the decrease in text metric scores becomes smaller as the availability of data declines. This might attribute to initial translations being low in quality.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
\textbf{Source} & \textbf{Reference} & \textbf{Translate} & \textbf{Refine\textsubscript{Contrast}} & \textbf{Paraphrase} \\
\hline
zh→en & According to a new decree, people must wear masks in indoor public places in Campania from now on, and offenders can be fined up to 1,000 euros. & A new regulation stipulates that in the Campania region, masks must be worn in indoor public places, with a maximum fine of 1,000 euros for those who violate the rule. & A new regulation states that in the Campania region, masks must be worn in indoor public places, with a maximum fine of 1,000 euros for those who violate the rule. & A new rule in Campania requires people to wear masks in indoor public places, and those who don’t follow this rule may be charged up to 1000 euros. \\
\hline
\end{tabular}
\caption{A Chinese-to-English example showing rich lexical choices and reduced translationese in the refined translation.}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
\textbf{Source} & \textbf{Reference} & \textbf{Translate} & \textbf{Refine\textsubscript{Contrast}} & \textbf{Paraphrase} \\
\hline
de→en & New regulation states that in the Campania region, masks must be worn in indoor public places, with a maximum fine of 1,000 euros for those who violate the rule. & The new regulation states that in the Campania region, masks must be worn in indoor public places, with a maximum fine of 1,000 euros for those who violate the rule. & A new regulation states that in the Campania region, masks must be worn in indoor public places, with a maximum fine of 1,000 euros for those who violate the rule. & A new rule in Campania requires people to wear masks in indoor public places, and those who don’t follow this rule may be charged up to 1000 euros. \\
\hline
zh→en & New regulation states that in the Campania region, masks must be worn in indoor public places, with a maximum fine of 1,000 euros for those who violate the rule. & The new regulation states that in the Campania region, masks must be worn in indoor public places, with a maximum fine of 1,000 euros for those who violate the rule. & A new regulation states that in the Campania region, masks must be worn in indoor public places, with a maximum fine of 1,000 euros for those who violate the rule. & A new rule in Campania requires people to wear masks in indoor public places, and those who don’t follow this rule may be charged up to 1000 euros. \\
\hline
en→zh & New regulation states that in the Campania region, masks must be worn in indoor public places, with a maximum fine of 1,000 euros for those who violate the rule. & The new regulation states that in the Campania region, masks must be worn in indoor public places, with a maximum fine of 1,000 euros for those who violate the rule. & A new regulation states that in the Campania region, masks must be worn in indoor public places, with a maximum fine of 1,000 euros for those who violate the rule. & A new rule in Campania requires people to wear masks in indoor public places, and those who don’t follow this rule may be charged up to 1000 euros. \\
\hline
\end{tabular}
\caption{Automatic scores of refining WMT 2021 news shared task German-to-English submissions.}
\end{table}
Table 5: Automatic scores of different strategies with GPT on low- and medium-resource pairs from WMT 2022 news translation.

<table>
<thead>
<tr>
<th></th>
<th>BLUE</th>
<th>chrF++</th>
<th>COMET_{DA}</th>
<th>COMET_{QE}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>en</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translate</td>
<td>23.00</td>
<td>25.89</td>
<td>.8863</td>
<td>.1255</td>
</tr>
<tr>
<td>Refine</td>
<td>22.63</td>
<td>27.30</td>
<td>.8941</td>
<td>.1305</td>
</tr>
<tr>
<td>Refine_{Contrast}</td>
<td>22.82</td>
<td>26.71</td>
<td>.8928</td>
<td>.1282</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>17.69</td>
<td>23.18</td>
<td>.8592</td>
<td>.1086</td>
</tr>
<tr>
<td><strong>fr</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translate</td>
<td>36.25</td>
<td>59.50</td>
<td>.8395</td>
<td>.0807</td>
</tr>
<tr>
<td>Refine</td>
<td>32.47</td>
<td>55.83</td>
<td>.8353</td>
<td>.0851</td>
</tr>
<tr>
<td>Refine_{Contrast}</td>
<td>33.12</td>
<td>56.37</td>
<td>.8308</td>
<td>.0805</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>16.06</td>
<td>44.28</td>
<td>.7937</td>
<td>.0682</td>
</tr>
</tbody>
</table>

4 Related Work

Closely related to translation refinement is automatic post-editing (APE), which trains a model to fix translation errors by learning from human correction data (Knight and Chander, 1994). While it has shown notable developments in statistical machine translation, it could be hindered in the deep learning era due to original translations being high-quality and lack of post-editing data (Junczys-Dowmunt and Grundkiewicz, 2018; Chatterjee et al., 2018). Whilst one way to facilitate this is more data provision (Chollampatt et al., 2020; Ive et al., 2020), our workaround utilizes LLM capabilities. Furthermore, a limitation of utilizing post-editing to alleviate translationese is that human editing data is collected from human annotators who are usually required to not make style improvements (Ive et al., 2020). Compared to post-editing, as an advantage, our method allows LLMs to re-generate an entirely different translation, which could escape the “post-editese” phenomenon (Toral, 2019).

There also exist approaches that do not rely on the source translation or human editing data (Simard et al., 2007). For instance, Freitag et al. (2019) trained a post-editor solely on monolingual data by reconstructing the original text given its round-trip translation. In our study, we incorporate stronger natural language modelling into post-editing by employing large language models. Other translation refinement research includes combining statistical and neural systems (Novak et al., 2016; Niehues et al., 2016), merging APE into the NMT framework (Pal et al., 2020; Chen et al., 2022), and debiasing translationese in the latent embedding space (Dutta Chowdhury et al., 2022). Typically translation error editing is a one-off process, whereas multi-round refinement can bring significant gains in the field of non-autoregressive translation decoding, where each output token is independent of other target positions (Lee et al., 2018; Gu et al., 2019; Xu and Carpuat, 2021).

With advanced language modelling and instruction following capabilities, LLMs have recently become highly effective tools for various NLP tasks (Radford et al., 2019; Brown et al., 2020; Chowdhery et al., 2022; Ouyang et al., 2022). Nowadays, it becomes infeasible yet unnecessary to optimize LLMs directly for specific tasks since they generalize to downstream tasks without explicit supervision. With more parameters and training data, LLMs may offer stronger performance than dedicated translation or post-editing models. The method we use to elicit a response from GPT is zero-shot hard prompting (Brown et al., 2020), which means affixing a description to the original task input to form a query to the model. Researchers have benchmarked LLMs’ capability to translate (Vilar et al., 2022; Zhang et al., 2023; Jiao et al., 2023; Hendy et al., 2023), and to evaluate translations (Kocmi and Federmann, 2023; Lu et al., 2023; Xu et al., 2023).

Recent findings show that GPT produces less literal translations, especially for out-of-English translations (Raunak et al., 2023a), which to some extent stands in contrast with our evaluation outcome. Concurrent with our study, Raunak et al. (2023b) formalized post-editing as a chain-of-thought process (Kojima et al., 2022) with GPT-4 and exhibited promising results. Different from their focus, our work features the iterative refinement process as a means to mitigate translationese. The improvement, especially for into-English, may be attributed to the abundant English pre-training data available for LLMs. To the best of our knowledge, although the concept of iterative refinement is not new, our study is the pioneering work in applying such strategies to LLMs for translation.
5 Conclusion

This work presented a simple strategy to translate and refine with large language models, in order to significantly reduce translationese in the outputs. We conclude that our method maintains translation quality and introduces lexical and structural changes, especially for high-resource into-English translation. Future work can explore open-source LLMs other than GPT. We also plan to expand human evaluation to larger samples and cover more languages.

6 Limitations

The term “translationese” is difficult to measure, so our human evaluation uses three proxies, namely “fluency”, “naturalness”, and “use of language”. While we adopted this approach following previous research, it might not eliminate biases such as an evaluator’s personal preference. The empirical results obtained in this paper are based on prompting GPT-3.5, and the conclusions need to be verified with other large language models.

Acknowledgement

We thank Nikolay Bogoychev and Biao Zhang for their feedback, as well as Laurie Burchell and Ziqin Fang for their help with human evaluation.

This work is funded by UK Research and Innovation (UKRI) under the UK government’s Horizon Europe funding guarantee (grant number 10052546).

References


